DOCUMENT RESUME

ED 079 642 CG 008 088

AUTHOR Neufeldt, Aldred H.
TITLE Considerations in the Implementation of Program Evaluation.
PUB DATE Mar 73
NOTE 27p.; Paper presented at the International Conference on Behavior Modification (5th Banff, Canada, March, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Contracts; *Cooperative Planning; *Methodology; Objectives; *Program Development; *Program Evaluation; Program Planning; *Research Problems; Technology; Values

ABSTRACT

Given the current, apparently favorable, climate for introducing and implementing program evaluation schema, the author questions factors mitigating against success, and variables relevant to program evaluation. He cites two types of problems: technological or instrumental problems of methodology and measurement, and evaluation process into technological versus value aspects and, in preparation for program evaluation, cites the need for the simultaneous consideration of both technological and value aspects. Besides the issues of incredibility and confidence, there are the following difficulties in program evaluation: (1) there is "no one way" to perform evaluation; (2) there is no generic logical structure that will assure a "right method of choice"; (3) evaluation ultimately becomes judgment as long as there is no ultimate ordering of priorities; and (4) the critical element in evaluation is who has the right to decide. Some specific suggestions for program evaluation conclude the report. (Author/LAA)

FILMED FROM BEST AVAILABLE COPY

ED 079642

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Considerations in the Implementation of Program Evaluations[1]

Aldred H. Neufeldt[2]

It has become quite fashionable in a variety of settings to talk about the "need" for program evaluation. Indeed, "program evaluation" has become an ostensible objective, if not reality, of program planners and implementers in health settings from general medicine through psychiatry, in educational settings from elementary school through university, in welfare settings, manpower training and placement settings, etc., etc., etc. Virtually every sphere of government activity today, along with many programs in the private domain, give evidence of this drive toward total evaluation of programs. In large part this drive is spurred by the anticipation of tremendous advantages (usually fiscal, though the altruist might anticipate social advantage as well) which might accrue as sound information is fed back into the planning, implementation and evaluation cycle.

Interest in evaluation has welled up relatively recently. It was no longer than five or ten years ago, at least in the settings I am familiar with, that the average conception of program evaluation involved a simple manipulation of even simpler statistics to gain yet simpler estimates of gain and loss. Today, a scant half decade later, the sophisticated health, education and social service program administrator talks expansively in terms of inputs and outputs, systems analyses, participative social action research strategies, multiple-regression analyses, and so on. One almost

CG 008 08

---

1. Paper presented at the Fifth Banff International Conference on Behavior Modification, March, 1973. The helpful comments by Mr. A.A. MacKinnon and Dr. J.C.K. Silzer are gratefully acknowledged.

2. Director of Operations Research, Psychiatric Services Branch, Saskatchewan Department of Health.

gets the impression that the day of planned and properly evaluated program change already has arrived. Alas, if this is our belief, I fear we are misled. It may be true that we are entering a society that increasingly is planned (controlled?). But, planned on the basis of what? While many gestures are made towards evaluation, it being a "good" word these days, thus far they appear to be relatively empty ones.

Certainly, it is difficult to find proper evaluations undertaken in practical settings. What about those that appear to be more academic? Several weeks ago, in a flush of enthusiasm, I decided to look at the number of annual publications of a program evaluation nature in several psychological journals felt to be particularly open to this type of publication. I was fully expecting, and hoping, to find an increase in such publications over the past decade. To my surprise, I could find no such increase. Indeed, according to my rather loose criteria, there was a relative decrease since the total number of articles published had increased while those of a program evaluation nature remained very small. Thus, I have concluded yet again, that much of the current outcry for program evaluation simply is the lip service of a social movement. In keeping with a typical social movement, there has been a rapid increase in public demand, the development of a special lingo, and a tremendous amount of enthusiasm. Like other social movements, I suspect that when it is found the goal of providing adequate program evaluation (nirvana?) cannot be easily achieved, that there will be a subsequent equally rapid fall-off in interest. From this point of view program evaluation currently is simply one of the hit parade "tunes" of today. Obviously, what we need to concern ourselves with, if we are serious about our intent to introduce program evaluation as something more than a transitory phenomenon, is that the content of the "tune" is sufficiently innovative to merit the label "classical".

Problem Sources

Given the current, apparently favorable, climate for introducing and implementing program evaluation schema, I find myself asking several questions such as the following: 'how is it that greater success has not been achieved to date?' 'What factors mitigate against success?' What are the relevant variables we need to consider when we implement a program evaluation?' These questions are posed on the assumption that if we can define the "what's" of the problem, it will be possible to move on to the "how's" of overcoming them. To properly consider them, we need to have a frame of reference. Let me briefly sketch several evaluative programs I either have participated in, or observed, to guide our thinking.

Example I

A new psychiatric centre was being planned for a region of our province that previously had been served only by a small out-patient clinic and by distant inpatient psychiatric facilities. The planning of this centre offered the opportunity of conducting a before-and-after study of the effect of opening such a centre on a previously grossly under-serviced region. As has been described elsewhere (Neufeldt, 1971; Neufeldt, Berrien and Smith, 1972), we worked out a neat quasi-experimental design, and innovated somewhat in the definition of concepts such as "subject" and "treatment conditions". Given the availability of an invaluable data source, in the form of a case and event-monitoring record system (Neufeldt, 1969), we were able to introduce a program evaluation schema with a minimum of fuss, and with reasonable success. There was a problem, however, that we were not readily able to get around. The problem, in that context, related to the question of what criteria do you use to evaluate the program. How do you measure program effectiveness? Indeed, how do you measure efficiency? In the

ultimate sense, efficacy would have to be determined in terms of treatment effectiveness on individual patients. However, as you know, the literature is replete with unsuccessful attempts at finding a broadly accepted definition of effective outcomes. Given that there was some need for haste to get our project off the ground (program evaluators almost always seem to become aware of changes somewhat late), and having concluded that we would not be able to improve on the previously unsuccessful attempts without some major input, yet having available to us patient-centered data, we decided to "cop out." We simply adopted the rationale used by Ullman (1967), limiting our criterion variables to those most characteristic of system functioning; i.e. admissions and discharges to inpatient care, the amount of outpatient contact, etc.

Example II

Over the years, beginning with the work of Ayllon (1962), we have had various experiences with introducing behavior modification programs. The introduction of these inevitably seems to give rise to a series of problems. The most recent such "behavior-mod" program was introduced in one of our major mental health settings; indeed, it still is in the process of being introduced!

Approximately one year ago, psychologists at the centre decided that a behavior modification scheme could profitably be introduced to assist in the rehabilitation of chronically hospitalized patients. Since such approaches have been fairly successful in the past, and since the methodology, definition of relevant variables, etc. is reasonably clear-cut, this seemed to be a relatively straight-forward task. A fair amount of background work was conducted, including the gaining of various types of approval from medical and non-medical staff. Commencing in the fall, concrete planning activities were undertaken, this including an in-service education program on the nature of behavior modifica-

tion programs. It looked as if "all systems were go." Then, just as subjects were about to be selected for admission into the program, resistances started to appear. As might have been anticipated, these resistances arose chiefly from the one particular professional group that increasingly has sensed itself as being in a beleaguered power position -- the medical personnel. To provide some brief background, this professional group, in Saskatchewan as elsewhere in North America, now finds that it's once sole authority over treatment is being diluted (some would say "violated") and, on the surface at least, also is gradually finding itself in a diminished administrative role. While this developing history has been welcomed by some within the profession, others have become quite defensive, as one might understand. In any event, some of the medical staff commenced blocking the introduction of the behavior-mod program. The immediate reason for blocking was perfectly legitimate. An elementary mistake had been made by the psychologists. They had failed to clear the selection process immediately prior to entering into the selection on the assumption that earlier permissions were sufficient. Nevertheless, this failure to seek the desired final approvals, was seized upon as the rationale for stopping a complete program that had been long in the planning. While it may yet get underway, the program currently is at a standstill.

Dimensions of Program Evaluation

These two examples outline the bare bones of the problem sources we face in implementing program evaluations. On the one hand, we have those that might be called "technological" and "instrumental" in nature -- problems of methodology and measurement, often held to be the essence of science; the "science of science." On the other hand, we have those which are very much of a human sort -- expressive components of human interaction, essentially

problems in the "art of science." In terms of the relative sophistication of the program evaluations, the second example given had the methodological aspects well taken care of. Sound methodologies were available, as were appropriate and relatively clear criterion variables. The problem most definitely lay in the expressive, or human dimensions -- the "art" of program implementation. This situation was reversed in the first example. No great problems were encountered from the human side; however, considerable problems of a technological-instrumental nature posed themselves -- specifically, the definition of criterion variables.

Conclusion number one, then, is that there are two important "elements" to be considered in undertaking program evaluations -- those that we might summarize with the term "instrumental", and those which might be termed "expressive." While all of us have some awareness of this dimension, a large portion of would-be researchers fail to distinguish clearly between the two. Because of the fact that our fellow human beings are going to have a not inconsiderable impact on any program evaluation undertaken, and given the perversity of those fellow human beings (program evaluators excepted, of course), it is absolutely vital that cautious and careful consideration be given to both Elements -- not merely those of an instrumental nature as most researchers are wont to do.

A second conclusion can be derived from the above, but usually only becomes clear when considering the meaning of the phrase "program evaluation." The key word, when we talk about program evaluation, is not so much the word "program" as the word "evaluation." Dictionaries variously define this word as "to place a value on", "find the amount of", or "to examine and judge." Throughout, the cue is the word <u>value.</u> Using this definition, published useages of the word evaluation, as well as the actual practices, can be divided into those in which "value" means:

(a) a number-value; or (b) a value-value. The former indicates some empirical worth, a number that has some absolute (or at least relative) worth. The latter implies a more subjective judgement -- that of preciousness, excellence, worth or desirability. If we examine the way in which any Element (in the sense described above) is selected for number-value measurement, it becomes apparent that such selection arises from the selector's priorities. That is, the number-value, in essence, is the outcome of a decision involving someone's value-value system. Thus, the program evaluation process can be dichotomized yet again into those aspects that to a large extent can be empirically determined and manipulated (hence, might be termed technological), and those aspects that are of a subjective, judgemental, value-value nature (hence personal value laden).

Division of what we have termed the Instrumental and Expressive Elements of evaluation into Technological and Value Aspects makes a good deal of inherent sense. We certainly need to concern ourselves with "technologies" such as experimental designs, criterion variables, strategies of implementation, etc. However, if our program evaluations are to be meaningful, and if they are to have some impact, then attention also needs to be given to considerations of a value-value nature. Four reasons might be advanced for this (cf. Weckworth, 1969). First, there is no one way to do evaluation. Second, there is no generic logical structure which will assure a unique "right method of choice." Third, evaluation ultimately becomes judgement and will remain so, so long as there is no ultimate criterion for monotonic ordering of priorities. And, fourth, the critical element in evaluation is simply: who has the right, i.e. the power, the influence or the authority, to decide.

Implications for Implementing Program Evaluations

Problems associated with the implementation of program evaluations, then, need to be considered from at least two dimensions -- those that have been termed Elements (Instrumental and Expressive), and those termed Aspects (Technological and Value). (A future paper will consider these two dimensions as they relate to a third -- "Macro" vs. "Micro" Level Problems). The remainder of this paper will offer a number of observations of the outcomes and implications associated with their cross-classification (see Figure 1).

Instrumental Elements

Instrumental elements typically are those to which greatest attention is given in academic training. For instance, courses in research design and statistics long have been a central part of graduate training in all of the social sciences. Despite this heavy emphasis, graduates have shown relatively little transfer of the learned skills from laboratory to practical settings. In part at least this is a function of the fact that most classes in methodology stereotypically have confined themselves, and the world of "science", to those mini-problems that could be dealt with in a neat way in laboratory sessions using introductory-psychology students or the pervasive white rodent. Obviously, the realm of program evaluation is not for those faculty and students whose personal dispositions allow for little deviation from equal-N, multivariate designs. Problems to be tackled within the program-evaluation sphere simply cannot always be fitted within those constraints. However, it should also be realized that there are some very real, seemingly intransigent problems apart from those of the limitations of would-be program evaluators.

Value Aspects. In a previous section reference has been made to the fact that personal value systems come into play in the very selection of problems for research, methodologies, criterion variables, and so on. It should be recognized, however, that even more fundamental value-related problems exist which tend to increase the difficulty of conducting evaluations.

The most commonly accepted operational definition of evaluation is: to compare accomplishment with stated objectives. This is a relatively straight forward definition, in that it is goal oriented. Since the operational definition is so simple, why then is evaluation so difficult? Some clue can be found if we look at the operational definition in some detail. In it, five assumptions are made:

(a) that the objectives are stated;

(b) in measurable terms;

(c) that accomplishments or outcomes are documentable;

(d) in the same measurable terms as the objectives; and

(e) that one knows what the word "compare" means, -- that is, we know what needs to be done both in the process of implementing and conducting the evaluation, as well as with the final observations.

Each one of these assumptions contains the seeds of difficulty. First, objectives very often are not stated. Too often no differentiation is made between the term "goal" as opposed to the term "objectives." Indeed, there seems to be a not inconsiderable degree of confusion; hence, little distinction between objectives and goals. Our goal may be to assess the efficacy of some new form of therapeutic intervention. However, this statement is insufficient in that it does not determine those intermediate states (objectives) that will allow for the goal to be reached. The goal of the Wright brothers was to fly. To

attain that goal, a wide range of intermediate objectives had to be reached -- the appropriate shape of wings to provide lott, the attainment of balance and control, the shape of propellors for proper power ratios, etc. While we may hear of purposes, of goals, of statements extolling motherhood and country, it regretfully is still a rare commodity to determine objectives. This simple step, in and of itself, would vastly increase both qualitative and quantitative successes in program evaluation.

Even if objectives are stated, many are not independent. In fact, they often are in conflict with one another and rarely would their summation add to the ultimate program goals. For example, the goal of our mental hospitals theoretically has been to cure people. Yet, many of the objectives inherent to operating the institution -- maintain patient and staff stability, encourage efficient operations, etc. -- have been inimical to other objectives more in concert with the goal.

In addition, the "state of the art" of evaluation is such that we have not developed the means of measuring most of our value system objectives. Unfortunately, there seems to be an inverse relationship between what is really important in life and what is easily measurable. Thus our measurement ineptness reflects both our ignorance and our errors.

Even if objectives can be stated, and appropriate measures made, very frequently there is difficulty in documenting accomplishment. In the even more rare event that accomplishment can be documented, commonly nobody knows what to do with it!. Or, if in fact someone knows, the comparison will still depend entirely on the judgement of whoever has the right to decide what to do with it. And so we start and conclude with "values" -- values in terms of the choice of goals and objectives; choice of measurements and ultimately choice of actions to be taken if and when outcomes are determined.

Technological Aspects. Having enumerated some of the problems that essentially are value-bound, what avenues are open to implementation? From the immediately preceding discourse the first line of attack should be obvious. Specify goals and objectives. The objectives should be operationally defineable, should be mutually exclusive, and should provide a logically additive means of achieving the ultimate goal. Second, devise a research strategem to encompass the various objectives specified. Third, select the necessary methodologies and criterion variables. Fourth, collect the necessary data, conduct appropriate analyses and determine implications of findings.

If these four steps sound familiar, they ought to since they form the essence of the "scientific method." The "scientific method" applies to program evaluation, as well as to other research endeavours. At the same time, it isn't as restrictive as some would have it. The "true" experimental design, with all variables except the critical independent variable controlled for, may be ideal; however, it is not the only legitimate approach to evaluation. The "scientific method" is not limited to this, or any other method. Quasi-experimental designs, while being vulnerable to certain internal threats (cf. Campbell and Stanley, 1966) and external threats (cf. Bracht and Glass, 1968) to validity, often are much more practicable and of greater utility. Indeed, our increase in knowledge has not come about so much from use of the so-called "true" experimental design as from the use of even more "primitive" methods of enquiry, all of which should be considered fair game in the conduct of evaluations.

At the same time, many seemingly intractable problems in fact, could legitimately be examined by means of "true" experimental and "quasi"-experimental designs. What often has prevented this from happening has been a misperception or, at

least, a limited perception of what a "subject" is (to use psychological jargon), and, therefore, what consequent "behaviors" legitimately are available for observation. In many settings the "subject" as traditionally defined, is most appropriate as the single unit of observation. That is, we observe the behaviors of individual organisms. However, this unit of observation can also be altered to be some combined behavior of a group of organisms -- as, for instance, in the circumstance where a new teaching technique is tried on a number of classrooms, there being equivalent classrooms receiving some control condition. In such a situation, entire classrooms can be treated as "subjects", or the unit of observation. Indeed, as has been discussed in some detail elsewhere (Neufeldt, Berrien and Smith, 1972), the unit of observation might be yet larger. Obviously, as the unit of observation changes, so do the possibilities of types of "behaviors" that might be observed.

Expressive Elements

While there exist a vast array of problems to be solved on the Instrumental side of the research coin, one at least is comforted by the feeling that they are potentially solveable. Typically, much less comfort is felt with respect to those problems essentially involving the perversity of our fellow man. Yet, this dimension too is potentially solveable. Certainly there are a number of program evaluators (alas, all too few) who have achieved a creditable track record at implementing program evaluations. Thus, there must be a certain technological aspect, as well as value aspect, to such successful implementations.

Value Aspects. Road-blocks, run-arounds and sabotage are not an uncommon part of the diet of program evaluators. In trying to determine the "Why's" of this fare, it would seem that

at least three value-laden aspects contribute -- credibility of the concept of program evaluation; confidence (or, lack of) in program evaluators, and the consequences associated with participating. Each will be considered in turn.

While endorsements of evaluation are given about as readily as a kind word for apple pie, the endorsers are not always so ready to open themselves to having their own programs evaluated. Thus, while program evaluations publicly are touted, in private they lack in credibility. As observed elsewhere, "no one enjoys being policed; moreover, most of us exhibit greater energy in contemplating the future than in mulling over the past. Thus, when evaluation is looked upon as little more than a search for past failures, it is not surprising that little enthusiasm is generated, especially among those whose wrists may get slapped in the process" (Reinke, 1972, p. 44). Such sceptical views of program evaluation and program evaluators are all too common. The air of the "efficiency expert" is still around. Regretfully, these attitudes are prevalent among administrators and planners, as well as among front-line workers.

In large part this would seem to be the result of a failure to understand two important roles that evaluation offers to the administrative planner: (1) to provide a means for continuing self-study, and (2) to offer periodic external review. Thus, "when the notion of continuing self-study is recognized to connote current and private surveillance designed to keep an individual or an agency in line with objectives -- or perhaps to help explain why this is impossible to do -- then the merits of evaluation begin to be appreciated. Further, when the review of accomplishments is designed for future improvement as well as past assessment, evaluation is likely to become still more palatable" (Reinke, 1972, p. 44). Reinke (1972) and others, suggest that evaluation procedures become more credible, not only to the extent that the

instrumental tools to be employed, enjoy reasonable validity and reliability (perhaps the single most frequent disclaimer used against implementation of evaluations), but also as the role of evaluation is perceived to alter from simply examining what "has been done, to what can be done." The ultimate goal and raison d' etre, then, is of utmost importance in establishing the credibility of program evaluations.

The second critical concern that the evaluator needs to give some thought to is the degree of confidence evaluatees have in him. Reference already has been made to the fact that a certain element of "fear" is aroused in evaluatees at the very mention of an evaluation. In part, at least, such lack of confidence is justified. Too often, in fact, evaluatees are instructed to participate in a so-called evaluation, only to find that the outcomes of the study either are non-sensical (hence, the participation a waste of time) or are used against them. In game-theory terms, evaluation has been experienced as a zero-sum game similar to others in which he was told to play, but where he had no chance to win. Obviously, some very real value decisions have to be made if an element of confidence or "trust" is to be established. In particular, careful thought has to be given to the type of contract that is established between evaluator and evaluatee.

Finally, value-decisions have to be made with respect to what actions will be undertaken consequent to completion of the evaluation, these decisions preferably being made prior to implementing the program evaluations. It is safe to say that planning without implementation leads to nothing, and that implementation without evaluation could conceivably make matters even worse. However, the degree to which evaluation findings are translated into revised plans and thence to improved performance is at least in part as much the responsibility of the evaluator, as it is of a programmer or implementer. In the process of achieving such a

translation, the very real political implications of evaluation and planning, as described by Campbell (1969) and Hall (1972), will need to be considered. A failure to do so leads to an emasculated form of evaluation, usually too late in conception and too light on consequence.

Technological Aspects. Since the Expressive Element is so central to the implementation and success of program evaluations, consideration will be given to ways (technologies) of meeting the value-related concerns parallel to the order in which they were raised in the previous section.

1. Credibility. Several tactics are of considerable importance to establishing credibility in the eyes of planners and administrators. The chief of these is that the ultimate goals, as well as the intermediate objectives, needs to relate directly to interests of the program leaders. Without such a direct and strong correlation, hopes of conducting an evaluation on a program of any significant degree of appropriateness or adequacy* are minimal. Once the interests of administrators have been determined, there is some utility in demonstrating that evaluators do indeed have and can fulfill an important service role to the administrators. In part this can be achieved through such devices as measuring achievement through existing standards and targets; however, additional attention should be given to whether original goals and objectives in fact were appropriate, whether resource development (facilities, manpower, etc.) is actually moving in the direction most suited to given conditions, whether the data gathering system is producing useful information, etc. Thus, credibility will be achieved if:

---

* "Appropriateness" might be defined as tackling a problem of some importance as opposed to one that is of mere parochial interest. "Adequacy" might be defined as whether a mountain is to be removed through the use of teaspoons or bulldozers.

(a) Evaluation is forward looking; and

(b) Evaluation exhibits concern for the relationship between ends and means.

2. Confidence. Confidence, or "trust", is somewhat of an ephemeral property that constantly needs bolstering in program leaders, as well as rank-and-file workers, for many of the reasons already cited. Given the fears that workers have of "zero-sum games", consideration should be given to renegotiating from an apparent zero-sum game to one that is non-zero-sum. That is, evaluatee as well as evaluator should legitimately be able to expect some positive and real profits or pay-offs (at least, no losses) to arise out of participation. In a recent, rather splendid article, Weinstein (1972) has critically examined the experimental games literature, as well as literature on source credibility in communication and research into inter-personal trust and friendship, distilling a number of findings on ways and means of establishing and maintaining such confidences. The technologies so determined, might be summarized as follows:

(a) From experimental-games research it has been found that co-operation between various parties, in games involving mixed motives (as in non-zero-sum games), improves when: each player is able to find a sensible pattern in the behavior of the other -- that is, that such responses of behaviors are predictable, and that they can be reliably anticipated; maximum co-operation only occurs when there are evident positive returns for all parties (mutual trust does not tend to develop in situations where pay-offs are so arranged that one party is strongly tempted to defect -- to act so as to harm the other); contracts are established to maintain a stable, co-operative degree of dependence (indeed, one might assume that contract negotiation process itself will help to build trust) and provide the necessary history of co-operation

for subsequent evaluation to be successful); and, communications between players are frequent, as well as open.

These findings lead to the following guideline for evaluation to succeed:

i. Spell out the rules of the game -- in particular, whether the "game" is zero-sum or non-zero-sum with respect to the evaluatee.*

ii. Negotiate a "contract", in Pratt's (1966) terms, in which the needs of both evaluator and evaluatee are clearly outlined, and in which the rules of action are summarized, preferably in written form. Allow for flexibility in revision of the contract.

iii. Use obviously confidential mechanisms for collecting data, since such a visible apparency has a reassuring effect on evaluatees that data will not be used against them individually.

iv. Provide an adequate amount of feedback on performance to evaluatees, usually presenting grouped data.

v. Establish the "humanness" of the evaluator through face-to-face discussions, using written communications for follow-up purposes only.

vi. Provide adequate opportunity for communication so that evaluatees are kept informed of every phase of the evaluation process.

vii. Make certain that there has been a history of co-operation before entering a high-threat area, such as feedback of performance, making certain that this perception is held by evaluatees as well as evaluators.

---

* "Evaluatee" is used in a broad sense, meaning the variety of "players" in the game from front-line workers to senior administrators.

b. Work on "source credibility" suggests that essentially five characteristics determine the credibility and trustworthiness of a speaker -- his perceived expertness, his reliability, his apparent intentions, his dynamism, and his personal attractiveness. The importance of a speaker's expertness will be readily apparent to anyone with any experience at all in negotiating the development of program evaluations. Technical knowledge of statistics and methodology of such evaluation are important, but not nearly sufficient. Indeed, if this is the only apparent knowledge of the evaluator, then he is likely to be readily dismissed as an "ivory tower ideologue." Consequently, it is vital that the evaluator either knows a good deal about the program being evaluated, as well as about program evaluation itself; or lets someone speak for him who does. The evaluator's intentions, as well as his reliability, will be conveyed through the delivery. One that is clear and logical is important. In addition, the choosing of a presenter who is perceived to be dynamic (i.e. has the quality of being more active than passive in the course of communication) seems to lead to greater trust. It may be, as Weinstein (1972) suggests, that such behavior indicates greater commitment by the speaker. Finally, the personal attractiveness of the speaker to the listener, while definitely an intangible, is vital. This suggests that trust might be especially difficult to develop when the evaluator represents a totally different milieu than that he wishes to evaluate.

From these observations, the following additional guidelines might be added.

viii. Choose a person who is knowledgeable about the program, as well as the instrumental technology of evaluation, to initiate and conduct the "negotiation" in establishing the program evaluation contract. If he already is a trusted person within the program, so much the better.

ix. Choose an active, dynamic negotiator, rather than one who is passive, since the former is likely to be more persuasive.

x. Do not choose a negotiator who is totally unattractive to evaluatees, this decision being based on criteria of attractiveness amongst evaluatees, not evaluators.

c. Finally, research into interpersonal trust also lends itself to certain conclusions. In particular, it provides some utility in determining the general level of trust that a group of "listeners" or "evaluatees" might have prior to approaching them. For instance, higher levels of trust tend to be found in individuals who feel less alienated and more in control of their lives, come from slightly higher socio-economic classes, etc. These, and other observations, provide the basis for several additional suggestions:

xi. Consider the personality and demographic characteristics of evaluatees before approaching them about the conduct of an evaluation. Extra caution will have to be used in those cases where individuals within a program are especially alienated.

xii. Consider the "trustworthiness" of the person introducing the notion of program evaluation, both in terms of the occupational group he is perceived to represent, and in terms of the group to be approached.

To these observations might be added the one that one of the most intractable obstacles to change is the innate human conviction that what everyone is used to doing must be right. Normal human pride of involvement leads to an almost uncontrollable subjective bias. Innovation requires both a willingness to give up even the most sacrosanct culturally accepted ways of doing

things and an openness to the new. Thus, the need for objectivity in evaluation should be self-evident. Not only must the evaluator guard against the subjective involvement that leads to selective misperceptions, he must also be trustworthy.

3. Consequences. Little will be said about the concern for what happens at the conclusion of a project, other than already has been said. From the previous observations, it will be obvious that this writer feels it imperative that an evaluator become involved in seeing to it that the implications of his findings are translated into recommendations of plan changes, and altered performances. This conclusion has been reached, partly because of the observation that research reports all too often become more efficient at gathering "dust" than of fingerprints; partly, because program planners frequently do not understand the technological jargon that tends to be inherent in evaluation reports; and, partly, because it will provide the evaluator with a better grasp of the practical considerations to be faced in implementing future evaluations and program changes.

Summary and Concluding Remarks

The implementation of program evaluations is by no means the easiest of tasks. The purpose of this paper has been to explore some of the issues related to this difficulty so as to more clearly ascertain the nature of these problems and to elucidate ways in which they might be overcome. The first observation was that research methodology, as traditionally taught in our university classes, typically has concentrated on the Instrumental Elements associated with such evaluation. Indeed, they usually have been restricted to those Instrumental Elements of a technological nature and ignored the value aspects. Observation two was that, while instrumental elements are important, an equally, and sometimes much more, important element is the Expressive one. That is, problems of human relations very often are as important as problems of methodology. Thus, the technological and value aspects of this dimension, too, were explored in some detail, it being felt quite possible to radically improve the way in which our evaluations are implemented. The sum and substance of the entire paper is that in the preparation for a program evaluation, both Elements have to be considered simultaneously. There is no sense working out a detailed research design when, in fact, it will not be acceptable to the people being evaluated. At the same time, if a good receptivity is achieved, it may be quite possible to develop a more rigorous and far-reaching research design than initially had been envisaged.

The above comments should in no way be used to underestimate the difficulties of implementation, however. In a sense, our knowledge of how to implement program evaluations, and the tools to use, is (to use an analogy* briefly invoked before) not unlike the state of knowledge available to the Wright brothers

---

*I am indebted to Dr. Carl Meilicke, Head, Health Services Administration, University of Alberta, for this observation.

when they ventured into the building of an aeroplane. A large number of very knowledgeable individuals from Leonardo da Vinci through Sir George Calley, Alexander Graham Bell, and Otto Lilienthal had the fond ambition to fly. Yet, they discovered that the problem of maintaining an object, heavier than air, in the air was of considerable complexity. Thus, the knowledge of physics and aerodynamics available to the Wright brothers was so crude that the development of the first successful aeroplane was, indeed, something of a marvel (cf. Fred C. Kelly's book Miracle at Kitty Hawk). By the time of the Wright brothers it had been learned that some source of power other than man should be used, and that the structure should have fixed rather than moveable wings. Virtually all of the remaining body of knowledge was yet to be discovered. Such basic insights that cambered wings provided greater upward lift than straight wings, that a vertical stabilizer was more efficient than a shift of body weight to maintain balance, and that such a stabilizer should have a moveable component, that certain types of propellors were more efficient than others, and, that rear, stabilizing wings were desireable, remained in the future. The Wright brothers were not alone, of course. Indeed, a large and hardy group of people were caught up by a wave of enthusiasm over flying, including builders and innovators, adventurers who wanted to fly. In addition, there were a large band of sceptics who said that "If God wanted us to fly he would have given us wings".

This analogy is not unlike the situation we face in program evaluation. We do have some knowledge as to what the ingredients of a successful evaluation are. In addition, there is a goodly amount of apparent information around which may, in fact, turn out to be misleading. Some program-situations exist where the independent variable to be manipulated is relatively simple, where the dependent variable is relatively straight

forward, and in which we can say we do some "program evaluation." However, in the main the programs we are dealing with are vastly complex. The independent variables to be manipulated have a number of levels of abstractions rather than simple single-level dimensions. Thus, selection of meaningful dependent variables becomes difficult as described above. These factors mean that, those of us who fancy ourselves to be "program evaluators", in reality should consider ourselves first of all to be builders and innovators of approaches to evaluation. Our main concern has to be to develop some basic knowledge as to what is possible, and what is appropriate. Comparable to the large band of sceptics who said that man would never fly, we have our share of sceptics as well. Frequently they take the form of administrators who view program evaluation as another punitive and threatening approach to their activity. Front line workers likewise find themselves in vulnerable positions. In addition, politicians while giving some lip service to the need for evaluation, usually proceed on their hap-hazard way of decision-making without seriously considering the potential that evaluative approaches might offer to the development of policy.

But the few short evaluative flights that have been made make me feel it quite possible and feasible for useful and effective evaluations to be undertaken. While we don't have detailed knowledge of all the relevant variables in our grasp as yet, we do at least have an idea as to the avenues of approach with the greatest probability ot success. These steps to effective action include the following.

1. Determine your objectives <u>and</u> goals. Both need to be determined, and with objectives derived that are in concert with the goals.
2. Plan the planning. Consider the various strategies of dealing with "people problems" as well as with "methodological problems".

3. Prepare the ground. Start advance work with people in the program to be evaluated so that the evaluation itself will be accepted.
4. Plan the evaluation. Make finalized evaluation plans in light of the preliminary acceptance of the idea among evaluatees.
5. Test the water. Despite the reluctance of most of us to waste time with "mere" pilot projects, the pilot-project route should be mandatory. It almost inevitably ends up saving much more time than would have been expended in undertaking the pilot work.
6. Let 'er fly! Having worked out bug-a-boo's as far as possible, swing into action, making sure that the evaluator considers all possible areas where things might go wrong, and that he has sufficient back-up support to meet all contingency problems as they arise.
7. Hope for serendipitous happenings. A little bit of luck never hurts!
8. Follow-through. Make certain that findings are lucidly translated into a language that program planners (from individual front-line practitioners to administrators) understand. Become an advocate of experimental change, keeping in mind both instrumental and expressive elements as well as technological and value aspects as described.

REFERENCES

AYLLON, T. and HAUGHTON, E. Control of the behavior of Schizophrenic patients by food. J. Exper. Anal. Behav., 1962, 5, 343-352.

BRACHT, G.H., and GLASS, G.V., The external validity of experiments. Amer. Ed. Res. J., 1968, 5, 437-474.

CAMPBELL, D.T. Reforms as experiments. Amer. Psychol., 1969, 24, 409-429.

CAMPBELL, D.T., and STANLEY, J.C. Experimental and Quasi-experimental Designs for Research. Chicago: Rand McNally and Co., 1966.

HALL, T.L. The political aspects of health planning. In, W.A. Reinke (Ed.), Health Planning: Qualitative Aspects and Quantitative Techniques. Baltimore: Waverly Press, Inc., 1972, 73-95.

NEUFELDT, A. H. A province-wide EDP system for community-based psychiatric services. Canad. Psychiat. Assoc. J., 1969, 14, 135-141.

NEUFELDT, A.H. Planning for comprehensive mental health programs. Canada's Mental Health, March-April, 1971, Monogr. Suppl. #67.

NEUFELDT, A.H., BERRIEN, V., and SMITH, L. The Developing Community Mental Health Centre: A study of referral and treatment patterns before and after the opening of a modern 'total care' psychiatric facility. Saskatoon: Marcotte Research Centre, Sask. Department of Health publ., 1972.

PRATT, S., and TOOLEY, J., Human Actualization Teams: the perspective of contract psychology. Amer. J. Orthopsychiat., 1966, XXXVI (5), 881-895.

REINKE, W.A. Methods and measurements in evaluation. In, W.A. Reinke (Ed.), Health Planning: Qualitative Aspects and Quantitative Techniques. Baltimore: Waverly Press, Inc., 1972, 44-52.

ULLMAN, L.P. Institution and Outcome: A Comparative Study of Psychiatric Hospitals. Toronto: Pergammon Press, 1967.

WECKWORTH, V.E. A conceptual model versus the real world of health care service delivery. In, C.E. Hopkins (Ed.), Outcomes Conference: Methodology of Identifying, Measuring and Evaluating Outcomes of Health Service Programs, Systems and Subsystems. Rockville, Maryland: Office of Scientific and Technical Information, NCHSRD, 1969, 39-62.

WEINSTEIN, M.S. The role of trust in program evaluation: some guidelines for the perplexed administrator. Canad. Psychol., 13, (3), 1972, 239-251.

Figure 1

"Element" and "Aspect" Dimensions of Program Evaluations

		Elements	
		Instrumental	Expressive
Aspects	Technological	Range of: - Research designs; - Operationally defined independent and dependent variables - Statistical techniques of analysis	Proven methods and techniques of: - introducing evaluation schema; - data collection; and, - ensuring future co-operation.
	Value	Considerations that lead to: - Choice of given criterion variables; - numbering systems chosen; - type of analyses conducted.	Considerations leading to: - decision to conduct program evaluation; - type of evaluation undertaken; - actions taken following evaluation.